IDENTIFIABILITY OF MIXTURES OF EXPONENTIAL FAMILIES

by

O. Barndorff-Nielsen

TECHNICAL REPORT NO. 38

February 17, 1964

PREPARED UNDER CONTRACT Nonr-225(28)

(NR-047-019)

OFFICE OF NAVAL RESEARCH

MATHEMATICS DEPARTMENT

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

IDENTIFIABILITY OF MIXTURES OF EXPONENTIAL FAMILIES

by

O. Barndorff-Nielsen

Let $\mathcal{F}_0 = \{F(\cdot|\tau) : \tau \in T\}$ be a family of n-dimensional distribution functions (d.f.s.) depending on a m-dimensional parameter τ which ranges over a Borel set T in R^m, the m-dimensional Euclidian space. We assume that for each fixed $x = (x_1, \dots, x_n) \in R^n$ the function $F(x|\cdot)$ is Borel measurable. Let $\mathcal{F}(\mathcal{G})$ denote the set of all probability measures (p.m.s.) on the Borel field $\mathcal{B}^n$ of R^n ($\mathcal{B}^m$ of R^m) and let $\mathcal{G}_T$ denote the set of those $\gamma \in \mathcal{G}$ for which $\gamma(T) = 1$. The family $\mathcal{F}_0$ determines a mapping $\psi : \mathcal{G}_T \to \mathcal{F}$ by the relation

$$(1) \qquad \psi(\gamma) = \int_T F(\cdot|\tau)\, d\gamma(\tau)$$

We speak of the d.f. $\psi(\gamma)$ as a mixture of $\mathcal{F}_0$ (w.r.t. γ). The mapping ψ is said to be identifiable if it is one to one. In certain connections (e.g. statistical estimation of γ) it is important to know whether ψ is identifiable. Various conditions for identifiability and nonidentifiability are known, see Teicher [4] and the references therein. Here we want to prove that, under mild restrictions, mixtures of exponential families $\mathcal{F}_0$ are identifiable. $\mathcal{F}_0$ is exponential (or of the Darmois-Koopman type) if for some σ-finite measure μ

$$(2) \qquad dF(x|\tau) = a(\tau)\, b(x)\, e^{\sum_{j=1}^{m} \tau_j h_j(x)}\, d\mu(x)$$

for $x \in R^n$, $\tau = (\tau_1, \tau_2, \ldots, \tau_m) \in T$, where $a(\tau) > 0$, $b(x) \geq 0$ and a, b, h_j, $j = 1, \ldots, m$ are all measurable.

Let $\gamma_1, \gamma_2 \in \mathcal{S}_T$ and let

$$(3) \qquad f_\nu(x) = \frac{d\psi(\gamma_\nu)}{d\mu} = b(x) \int_T a(\tau)\, e^{\sum_1^m \tau_j h_j(x)}\, d\gamma_\nu(\tau), \qquad \nu = 1, 2.$$

Furthermore, let $\xi = \{x : f_1(x) = f_2(x) \neq 0\}$, let $\eta = \{y = (h_1(x), \ldots, h_m(x)) : x \in \xi\}$ and let

$$(4) \qquad f_\nu^*(y) = \int_T a(\tau)\, e^{(\tau, y)}\, d\gamma_\nu(\tau), \qquad \nu = 1, 2.$$

where (τ, y) denotes the inner product of $\tau (\in T)$ and $y \in R^m$. Then $f_1^*(y) = f_2^*(y)$ if $y \in \eta$; our aim is to show that under certain further restrictions this implies $\gamma_1 = \gamma_2$. Let $c(\eta)$ denote the convex hull of η. We shall distinguish between four cases.

(i) η is finite.

(ii) η is infinite, $c(\eta)$ is bounded and η does not have an accumulation point in the interior of $c(\eta)$.

(iii) As (ii) except that $c(\eta)$ is assumed unbounded.

(iv) η is infinite and η has an accumulation point in the interior of $c(\eta)$.

Case (i). The important example of this case is the binomial distribution. An analysis of the identifiability problem for that distribution can be found in [4].

Case (ii). From the viewpoint of statistics (ii) is the case of least interest. We have obtained no general results. The problem is essentially this: $(n = m = 1)$. Let γ_1 and γ_2 be two p.m.'s on $(R, \mathcal{B})$ whose Laplace transforms $\varphi_1(z)$ and $\varphi_2(z)$ both exist in a strip $0 \leq \mathcal{R}e\ z \leq \rho$, $\rho > 0$. Let $\{x_n\}$ be a sequence of real numbers such that $0 < x_n \leq \rho$ for all n and $x_n \to 0$ as $n \to \infty$. Find conditions under which $\varphi_1(x_n) = \varphi_2(x_n)$ for all n implies $\varphi_1(it) = \varphi_2(it)$ for all real t (i.e., identity of the Fourier transforms of γ_1 and γ_2 and hence identity of γ_1 and γ_2).

Case (iii). We shall treat the subcase:

(iii)'. η contains the set I^+ of all lattice points in R^m with nonnegative components, i.e., $I^+ = \{k = (k_1, \ldots, k_m) : k_j$ is a nonnegative integer, $j = 1, \ldots, m\}$.

We have, since $0 = (0, \ldots, 0) \in I^+$

$$f_1^*(0) = \int_T a(\tau)\ d\gamma_1(\tau) = \int_T a(\tau)\ d\gamma_2(\tau) = f_2^*(0). \tag{5}$$

Let us denote the common (positive) value in (5) by c and let us introduce the p.m.'s γ_ν^*, $\nu = 1, 2$, by $d\gamma_\nu^*(\tau) = c^{-1} a(\tau)\, d\gamma_\nu(\tau)$. Thus

$$(6) \quad f_1^*(k) = \int_T e^{(\tau,k)}\, d\gamma_1^*(\tau) = \int_T e^{(\tau,k)} d\gamma_2^*(\tau) = f_2^*(k) \qquad \forall\, k \in I^+ .$$

Let w be the transformation : $\tau \to \lambda = w(\tau)$ where $\lambda = (\lambda_1, \ldots, \lambda_m) = (e^{\tau_1}, \ldots, e^{\tau_m})$; let $\Lambda = w(T)$ and $\pi_\nu = \gamma_\nu^* w^{-1}$, $\nu = 1, 2$. We obtain from (6)

$$(7) \quad \mu_{k_1 \cdots k_m} = \int_\Lambda \lambda_1^{k_1} \cdots \lambda_m^{k_m}\, d\pi_1(\lambda) = \int_\Lambda \lambda_1^{k_1} \cdots \lambda_m^{k_m}\, d\pi_2(\lambda)$$

$$\forall\, k = (k_1, \ldots, k_m) \in I^+.$$

We can draw the following conclusion.

Proposition 1. Suppose that assumption (iii)' is satisfied and suppose that π_1 and π_2 are uniquely determined by their moments (7). Then $\pi_1 = \pi_2$ and consequently $\gamma_1 = \gamma_2$.

In order to derive a sufficient condition for $\gamma_1 = \gamma_2$ which is more useful than that of Proposition 1 we state the following lemma.

Lemma 1. Let π be an arbitrary p.m. on $(R^m, \mathcal{B}^m)$ with $\pi(R^{+m}) = 1$ where R^+ is the set of nonnegative reals and with all moments

$$(8) \qquad \mu_{k_1 \cdots k_m} = \int_{R^m} \lambda_1^{k_1} \cdots \lambda_m^{k_m}\, d\pi(\lambda), \qquad k \in I^+$$

finite. If there exists a positive number ρ such that the series

$$(9) \qquad \sum_{k \in I^+} \mu_{k_1 \cdots k_m} \frac{\rho^{k_1 + \cdots + k_m}}{k_1! \cdots k_m!}$$

is convergent then π is the unique p.m. with these moments.

The lemma and its proof are straightforward generalizations of a result in the book of Cramer [2; 176].

Let us apply the lemma to (7). We find (dropping the subscript ν)

$$0 \leq \sum_k \mu_{k_1 \cdots k_m} \frac{\rho^{k_1 + \cdots + k_m}}{k_1! \cdots k_m!}$$

$$= \int_\Lambda \sum_k \prod_{j=1}^m \frac{(\lambda_j \rho)^{k_j}}{k_j!} \, d\pi$$

$$= \int_\Lambda \prod_{j=1}^m \left(\sum_{\ell=1}^\infty \frac{(\lambda_j \rho)^\ell}{\ell!} \right) d\pi$$

$$= \frac{1}{c} \int_T a(\tau) \, e^{\rho \Sigma e^{\tau_j}} \, d\gamma(\tau)$$

$$\leq \frac{1}{c} \sup_{\tau \in T} a(\tau) \, e^{\rho \Sigma e^{\tau_j}} .$$

Therefore

Proposition 2. Suppose that assumption (iii)' is satisfied and suppose that

$$\text{(10)} \qquad \sup_{\tau \in T} a(\tau)\, e^{\rho \Sigma e^{\tau_j}} < \infty$$

for some $\rho > 0$. Then $\gamma_1 = \gamma_2$.

As an application, let us consider the instance where $n = m$, $h_j(x) = x_j$ (j-th coordinate of x; $j = 1, 2, \ldots, m$) and where the measure μ in (2) is concentrated on I^+; then without loss of generality we can and will assume μ to be counting measure on I^+. Hence the family $\mathcal{F}_0$ is given by

$$\text{(11)} \qquad F(x|\tau) = \begin{cases} \sum_{k=0}^{[x]} a(\tau)\, b(k)\, e^{(\tau,k)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

in an obvious notation. Assumption (iii)' becomes: $b(k) > 0\ \forall k \in I^+$ and we have

Corollary 1. If the family $\mathcal{F}_0$ given by (11) satisfies $b(k) > 0\ \forall k \in I^+$ and

$$\text{(12)} \qquad \sup_{\tau \in T} a(\tau)\, e^{\rho \Sigma e^{\tau_j}} < \infty$$

for some $\rho > 0$ then ψ is identifiable.

Specializing still further we obtain (Feller [3])

Corollary 2. The mapping ψ determined by the Poisson family $\mathcal{F}_0 = \{F(\cdot|\tau) : -\infty < \tau < \infty\}$, where

$$F(x|\tau) = \sum_{k=0}^{[x]} e^{-\lambda} \frac{\lambda^k}{k!}, \qquad x \geq 0, \quad \lambda = e^{\tau}$$

is identifiable.

Case (iv). We shall prove that $\gamma_1 = \gamma_2$ provided

(iv)'. There exists an accumulation point $y^{(o)} = (y_1^{(o)}, \ldots, y_m^{(o)})$ of η in the interior of $c(\eta)$ with the following property. If two arbitrary complex power series

$$\Sigma\, a^{(\nu)}_{j_1 j_2 \cdots j_m} (z_1 - y_1^{(o)})^{j_1} (z_2 - y_2^{(o)})^{j_2} \cdots (z_m - y_m^{(o)})^{j_m}, \qquad \nu = 1, 2$$

coincide for all $z = (z_1, \ldots, z_m) \in \eta \cap V$ for some neighborhood V of $y^{(o)}$, then they have identical coefficients.

We note that assumption (iv)' is equal to (iv) if $m = 1$. A sufficient condition for (iv)' is that η be dense in some open subset of R^m.

Proposition 3. Suppose that assumption (iv)' is satisfied. Then $\gamma_1 = \gamma_2$.

Proof. Without loss of generality we can and will assume that the origin 0 is in η and that there is a neighborhood $K = \{y : |y_j| < \rho,\ j = 1, \ldots, m\}$ of 0 for which $K \subset c(\eta)$ and K contains $y^{(o)}$. Then

$$(13) \qquad f_1^*(0) = \int_T a(\tau)\ d\gamma_1(\tau) = \int_T a(\tau)\ d\gamma_2(\tau) = f_2^*(0) .$$

Let us denote the common value in (13) by c and let us define the p.m.s. γ_ν^*, $\nu = 1, 2$ by $d\gamma_\nu^*(\tau) = \frac{1}{c}\, a(\tau)\ d\gamma_\nu(\tau)$. Furthermore, let φ_ν, $\nu = 1, 2$ denote the Laplace transform of γ_ν

$$\varphi_\nu(z) = \int_T e^{(\tau, z)}\ d\gamma_\nu^*(\tau)$$

where $z = (z_1, \ldots, z_m)$, $z_j = u_j + iv_j$ $(j = 1, \ldots, m)$. φ_ν exists for all $z \in K' = \{z | u = (u_1, \ldots, u_m) \in K\}$. In fact, for any such z, $|\exp((\tau, z))| \leq \exp((\tau, u))$ and a moments reflection shows that there exists a $y \in \eta$ with $(\tau, u) \leq (\tau, y)$; thus

$$\int_T |e^{(\tau, z)}|\, d\gamma_\nu^*(\tau) \leq \frac{1}{c} \int_T a(\tau)\ e^{(\tau, y)}\ d\gamma_\nu(\tau) < \infty .$$

More is true: φ_ν is an analytic function of $z = (z_1, \ldots, z_m)$ in the domain K'. To prove this it suffices to show that φ_ν is analytic in each of the variables z_j, $j = 1, \ldots, m$ (see [1]). Hence let us consider

$$(14) \qquad \frac{\varphi_\nu(z + he_j) - \varphi_\nu(z)}{h} = \int_T e^{(\tau,z)} \, \frac{e^{\tau_j h} - 1}{h} \, d\gamma^*_\nu(\tau)$$

where $z = u + iv \in K'$, e_j denotes the j-th unit vector in R^m and h is an arbitrary conplex number. Let $\delta > 0$ be so small that $z + he_j \in K'$ for all h such that $|h| \leq \delta$. Using the (well known) inequality

$$\left|\frac{e^{\tau_j h} - 1}{h}\right| \leq \frac{e^{|\tau_j|\delta}}{\delta} \qquad \text{for} \quad |h| \leq \delta$$

we find that the integrand in (14) is dominated by

$$\frac{1}{\delta}\left(e^{(\tau,\ u + \delta\ e_j)} + e^{(\tau,\ u - \delta\ e_j)} \right)$$

and since the integral of this quantity is finite we may pass to the limit $h \to 0$ under the integration sign in (14) to obtain

$$\frac{\varphi(z + he_j) - \varphi(z)}{h} \to \int_T \tau_j \, e^{(\tau,z)} \, d\gamma^*_\nu(\tau) \qquad \text{as} \quad h \to 0 \, .$$

We have thus shown that φ_ν is analytic in K'. Consequently φ_ν can be expanded in a power series around $z^{(o)} = y^{(o)}$

$$\varphi_\nu(z) = \Sigma \, a^{(\nu)}_{j_1 j_2 \cdots j_m} (z_1 - y_1^{(o)})^{j_1} (z_2 - y_2^{(o)})^{j_2} \cdots (z_m - y_m^{(o)})^{j_m}$$

the expansion being valid in some neighborhood V of $y^{(o)}$. We have $\varphi_1(z) = \varphi_2(z)\ \forall\ z \in \eta$ and hence, by assumption (iv)' and uniqueness of analytic continuation, $\varphi_1(z) = \varphi_2(z)\ \forall\ z \in K'$. In particular $\varphi_1(z) = \varphi_2(z)$ for all purely imaginary $z = iv = (iv_1, \ldots, iv_m)$, i.e., the characteristic functions of γ_1^* and γ_2^* coincide, hence $\gamma_1^* = \gamma_2^*$ or, equivalently, $\gamma_1 = \gamma_2$. q.e.d.

By the remark preceding Proposition 3, we obtain

<u>Corollary 3</u>. Suppose that in the representation (2): (a) μ is n-dimensional Lebesgue measure, (b) the functions h_j, $j = 1, \ldots, m$ are all continuous, (c) the set $\{y : y = (h_1(x), \ldots, h_m(x)), b(x) > 0, x \in R^n\}$ contains a (nonempty) open set. Then ψ is identifiable.

Specializing still further we get

<u>Corollary 4</u>. Suppose that $\mathcal{F}_0$ is the Gaussian family

$$\mathcal{F}_0 = \{F(\cdot|\tau)|\tau = (\tau_1, \tau_2), -\infty < \tau_1 < \infty, 0 < \tau_2 < \infty\},$$

$$(15) \qquad \frac{dF(x_1, \ldots, x_n|\tau_1, \tau_2)}{d\mu} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{\ell=1}^{n} (x_\ell - \xi)^2\right)$$
$$= \left(\frac{\tau_2}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{\tau_1^2}{\tau_2}\right) e^{h_1(x)\tau_1 + h_2(x)\tau_2}$$

where μ is n-dimensional Lebesgue measure, $\tau_1 = \xi\sigma^{-2}$, $\tau_2 = \sigma^{-2}$, $h_1(x) = \Sigma x_\ell$ and $h_2(x) = -\frac{1}{2}\Sigma x_\ell^2$. If $n > 1$, then ψ is identifiable (Teicher has shown, see [5], that ψ is not identifiable if $n = 1$).

REFERENCES

[1] Bochner, S. and Martin, W.T., "Several complex variables," Princeton University Press (1948).

[2] Cramér, H., "Mathematical methods of statistics," Princeton University Press (1957).

[3] Feller, W., On a general class of contagious distributions, Ann. Math. Stat. 14 (1943) 389-399.

[4] Teicher, H., Identifiability of finite mixtures, Ann. Math. Stat. 34 (1963) 1265-1269.

[5] Teicher, H., On the mixture of distributions, Ann. Math. Stat. 31 (1960) 55-73.